Published in final edited form as:
Biochemistry. 2013 December 3; 52(48): 8663–8676. doi:10.1021/bi401192z.

Structural and biochemical characterization of the bilin lyase CpcS from *Thermosynechococcus elongatus*

Christina M. Kronfel[1], **Alexandre P. Kuzin**[2], **Farhad Forouhar**[2], **Avijit Biswas**[1], **Min Su**[2], **Scott Lew**[2], **Jayaraman Seetharaman**[2], **Rong Xiao**[3], **John K. Everett**[3], **Li-Chung Ma**[3], **Thomas B. Acton**[3], **Gaetano T. Montelione**[3], **John F. Hunt**[2], **Corry E. C. Paul**[1], **Tierna M. Dragomani**[1], **M. Nazim Boutaghou**[4], **Richard B. Cole**[4,5], **Christian Riml**[1], **Richard M. Alvey**[6], **Donald A. Bryant**[6,7], and **Wendy M. Schluchter**[1,*]

[1]Department of Biological Sciences, University of New Orleans, New Orleans, LA 70148 USA

[2]Department of Biological Sciences, Northeast Structural Genomics Consortium, Columbia University, New York, NY 10027 USA

[3]Center for Advanced Biotechnology and Medicine, Department of Molecular Biology and Biochemistry, and Department of Biochemistry, Robert Wood Johnson Medical School, and Northeast Structural Genomics Consortium, Rutgers, The State University of New Jersey, Piscataway, NJ 08854 USA

[4]Department of Chemistry, University of New Orleans, New Orleans, LA 70148 USA

[5]Institut Parisien de Chimie Moléculaire, UMR 7201, Université Pierre et Marie Curie (Paris 6), 4 Place Jussieu, 75252 Paris, France

[6]Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA 16802 USA

[7]Department of Chemistry and Biochemistry, Montana State University, Bozeman, MT 59717 USA

Abstract

Cyanobacterial phycobiliproteins have evolved to capture light energy over most of the visible spectrum due to their bilin chromophores, which are linear tetrapyrroles that have been covalently attached by enzymes called bilin lyases. We report here the crystal structure of a bilin lyase of the CpcS family from *Thermosynechococcus elongatus* (*Te*CpcS-III). *Te*CpcS-III is a 10-stranded beta barrel with two alpha helices and belongs to the lipocalin structural family. *Te*CpcS-III catalyzes both cognate as well as non-cognate bilin attachment to a variety of phycobiliprotein subunits. *Te*CpcS-III ligates phycocyanobilin, phycoerythrobilin and phytochromobilin to the alpha and beta subunits of allophycocyanin and to the beta subunit of phycocyanin at the Cys82-equivalent position in all cases. The active form of *Te*CpcS-III is a dimer, which is consistent with the structure observed in the crystal. Using the UnaG protein and its association with bilirubin as a

*Corresponding Author: Mailing Address: Dr. Wendy M. Schluchter, Department of Biological Sciences, University of New Orleans, 2000 Lakeshore Drive, New Orleans, LA 70148. Phone (504) 280-7194; fax (504) 280-6121; wschluch@uno.edu.

The Protein Database ID for *Te*CpcS is 3BDR.

Conflict of Interest Statement:
The authors declare no competing financial interests.

Supporting information available:
One table and seven figures are included as supplementary information for this manuscript. This material is available free of charge via the Internet at http://pubs.acs.org.

guide, a model for the association between the native substrate, phycocyanobilin, and *Te*CpcS was produced.

Keywords

bilin lyase; cyanobacteria; fluorescent probes; phycobiliproteins; lipocalins

Light harvesting in cyanobacteria is accomplished by phycobilisomes, supramolecular structures principally comprised of phycobiliproteins (PBPs) with covalently attached chromophores [1]. The chromophores, termed bilins, arise from the oxidative cleavage of heme [2–5]. Biliverdin IXα is subsequently reduced by ferredoxin-dependent reductases to form phycocyanobilin (PCB), a blue-colored chromophore present in all cyanobacteria and phycoerythrobilin (PEB), a red-colored chromophore present in some but not all cyanobacteria [6, 7]. In plants, biliverdin IXα is reduced by the enzyme HY2 to form phytochromobilin (PΦB) [6], a blue-green bilin which attaches to the light-sensor phytochrome [8].

Bilins have energy transfer properties that are greatly enhanced when their conformations are extended and made more rigid by their association with an appropriate PBP [1, 9]. Bilins are covalently attached via thioether bonds to conserved cysteine residues on the alpha and beta subunits of the PBPs [10]. A large linker PBP, which serves as the scaffold for phycobilisome core assembly and which anchors the phycobilisome to the thylakoid membrane, is the only protein in the cyanobacterial phycobilisome that can self-ligate its PCB chromophore [11–13]. All other PBPs require bilin lyase enzymes for ligation of their bilin chromophores [14–20]. Bilin lyases are loosely classified according to the substrate PBP subunit (either α or β), the Cys residue to which the bilin is attached, and their amino acid sequences [17, 21]. Although their specificity for bilin and PBP substrates appears to be highly specific *in situ*, heterologous expression studies in *Escherichia coli* have revealed that bilin lyases can ligate diverse bilins to an individual PBP subunit [13, 17, 19, 22–24].

The first bilin lyase to be described in detail was the CpcE/CpcF lyase of *Synechococcus* sp. PCC 7002. This heterodimeric lyase is representative of the CpcE/CpcF lyase family and attaches PCB to CpcA (α-phycocyanin) at Cys84 [15, 20, 25–27]. PecE/PecF and CpeY/CpeZ, which are paralogs of CpcE/CpcF, are responsible for bilin ligation on the analogous cysteine residues of α-subunits of other rod proteins (PecA or α-phycoerythrocyanin and CpeA or α-phycoerythrin, respectively) [28–30]. It is postulated that the bilin lyase/isomerases RpcG and MpeZ arose as a fusion of genes encoding a CpcE/CpcF-type lyase [23, 31], and some initial modeling of MpeZ suggests that it adopts a structure that contains primarily α-helices [31].

Members of the two other lyase families, denoted the CpcT and the CpcS/CpcU families, were also initially identified and characterized from *Synechococcus* sp. PCC 7002 [16, 18, 32, 33]. The CpcT bilin lyase attaches PCB at Cys153 of CpcB and PecB [18, 34] and members of this family are probably distantly related to those present in the CpcS/CpcU family. There are members of this CpcS/CpcU family that ligate PCB to CpcB or PecB (at Cys82 position) and to the α and β subunits of allophycocyanin (AP) [13, 16, 19, 33]; these members are typically given the designation of "CpcS" or "CpcU". Other members of this group are given the designation "C<u>pe</u>S" or "C<u>pe</u>U" because they are encoded by genes which cluster together in operons with genes that encode <u>p</u>hyco<u>e</u>rythrin subunits [33], and some of these CpeS-type lyases have been shown to attach PEB to CpeB (β-phycoerythrin) at Cys80 [28, 35]. A heterodimer composed of CpcS-I and CpcU targets Cys82 of CpcB (β-phycocyanin) and the equivalent Cys on both subunits of AP [16]. In some cyanobacterial species the S-type lyase is a homodimer [35] or it functions as a monomer [17, 36].

Here we report the crystal structure of CpcS from *Thermosynechococcus elongatus* BP-1 (*Te*CpcS), which was determined as part of a structural genomics initiative [37, 38]. *T. elongatus* does not synthesize PEB and contains only PC and AP as major PBPs [39]. Because of strong sequence similarity among *Te*CpcS, CpcS-I and CpcU from *Synechococcus* sp. PCC 7002, and CpcS of *Nostoc* sp. PCC 7120, *Te*CpcS was postulated to be a bilin lyase. According to the CyanoLyase database (http://cyanolyase.genouest.org/), both *Te*CpcS and CpcS from *Nostoc* sp. PCC 7120 belong to the CpcS-III subfamily [40]. Using a heterologous plasmid co-expression system in *E. coli*, we establish here that *Te*CpcS protein is a bilin lyase. We show that *Te*CpcS is a homodimer, has highly flexible substrate specificity and can attach any of three bilin substrates, PEB, PCB and PΦB, to suitable apo-PBP substrates in *E. coli*. *Te*CpcS was able to attach bilin chromophores to Cys82-equivalent position of ApcA, ApcB, and CpcB but was unable to ligate any bilin to CpcA.

MATERIALS AND METHODS

Construction of expression vectors

Most of the plasmids used in this study have previously been described and are listed in Table 1. The *cpcS* gene from *T. elongatus* (accession number Q8DI91) was amplified by polymerase chain reaction (PCR) using primers TEcpcSF (5′-tccccattagCATATGtgcataggtatggacatccgc-3′, added NdeI site in capital letters) and TEcpcSR (5′-gaaaaaCTCGAGggagttggcgggttgcgtc-3′, added XhoI site in capital letters), digested with NdeI and XhoI, and ligated into similarly digested pCOLADuet-1 (Novagen, Madison, WI). Recombinant *Te*CpcS expressed from this plasmid contains a C-terminal S-tag. For crystallization, mass spectrometry and size-exclusion experiments, the *cpcS* gene was cloned in pET21c to create the clone pTER13-21 with a C-terminal His-tag (described below). For mutagenesis, the *cpcS* gene was also cloned in a similar way in pET30c to create pTER13-30. The pTER13-30 was also used to produce site-specific variants using standard methods as described[28] and using the following primers: *Te*CpcS(R151G): 5′-[Phos]ccccaatttaggtctgcgcacca-3′; *Te*CpcS(S155G): 5′-[Phos]tctgcgcaccggtattctcaagc-3′; *Te*CpcS(C169S): 5′-[Phos]ggcctccttctcctcggaaattcg-3′; pET30c(XbaI)del: 5′-[Phos]gtgagcggataacaattcccctctacaaataattttg-3′; *Te*CpcS(C2S).F.Nde: 5′-gatataCATATGtccataggtatggacatccgcg-3′; *Te*CpcS(C2S).R.Xho: 5′-ggtgCTCGAGggagttggcgggttg-3′. The five plasmids produced were pTER13(R151G), pTER13(S155G), pTER13(C2S), pTER13(C169S), and pTER13(C2S/C169S) (see Table 1). All constructs were verified by standard DNA sequence analyses.

Protein expression and purification for crystallization

The production of *Te*CpcS-III protein was carried out as part of the high-throughput protein production process of the Northeast Structural Genomics Consortium (NESG)[41, 42]. *Te*CpcS-III corresponds to NESG Target TeR13. The full-length *cpcS-III* (*ycf58*) gene from *Thermosynechococcus elongatus* (strain BP-1) was cloned into a pET21 (Novagen) derivative, generating plasmid pTER13-21. The resulting recombinant protein contains eight non-native residues (LEHHHHHH) at the C-terminus. *E. coli* BL21 (DE3) pMGK cells, a rare codon enhanced strain, were transformed with pTER13-21. A single isolate was cultured in MJ9 minimal media [43] supplemented with selenomethionine, lysine, phenylalanine, threonine, isoleucine, leucine and valine for the production of selenomethionine-labeled *Te*CpcS-III [44]. Initial growth was carried out at 37 °C until the $OD_{600\ nm}$ of the culture reached 0.6–0.8. The incubation temperature was then decreased to 17 °C, and protein expression was induced by the addition of isopropyl-β-D-thiogalactopyranoside (IPTG) at a final concentration of 1 mM. Following overnight incubation, the cells were harvested by centrifugation.

Selenomethionyl *Te*CpcS-III was purified by standard methods. Cell pellets were resuspended in lysis buffer (50 mM NaH_2PO_4 (pH 8.0), 300 mM NaCl, 10 mM imidazole and 5 mM β-mercaptoethanol) and disrupted by sonication. The resulting lysate was clarified by centrifugation at 26,000 ×*g* for 45 min at 4 °C. The supernatant was loaded onto a nickel nitrolotriacetic acid (Ni-NTA) column (Qiagen, Inc, Chatsworth, CA) and eluted in lysis buffer containing 250 mM imidazole. Fractions containing partially purified *Te*CpcS-III were pooled and buffer conditions providing monomeric samples were optimized by analytical gel filtration detected by static light scattering, following the protocol described elsewhere[41, 42]. Preparative gel filtration (Superdex 75, GE Healthcare, Piscataway, NJ) was then performed using a buffer containing 10 mM Tris-HCl (pH 7.5), 100 mM NaCl, 5 mM dithiothreitol (DTT), and 0.02% NaN_3. The purified *Te*CpcS-III protein was concentrated to 7.6 mg/ml, flash frozen in aliquots, and used for crystallization screening. Sample purity (>97%) and molecular mass (21.914 kDa) were verified by SDS-PAGE and MALDI-TOF mass spectrometry, respectively. The yield of purified protein was approximately 5 mg/L.

Protein crystallization

The *Te*CpcS-III free enzyme was crystallized at 20 °C by the hanging-drop vapor diffusion method. For the free enzyme crystals, aliquots (2 μl) of protein solution containing *Te*CpcS-III (7.6 mg/ml in 10 mM Tris-HCl (pH 7.5), 100 mM NaCl, 5 mM DTT, 0.02% NaN_3) was mixed with an equal volume of the reservoir solution consisting of 0.1 M *N*-cyclohexyl-3-aminopropanesulfonic acid, (CAPS, pH 10.0), 18% polyethylene glycol 20000, and 0.1M NH_4H_2PO4. The crystals were cryo-protected by transferring them to a crystallization solution that was supplemented by 20% (v/v) glycerol. The crystals were flash-frozen in liquid propane for data collection at 100 K.

The free enzyme crystals belong to space group $P4_12_12$, with cell parameters of a=b=75.01 Å and c=83.36 Å There is one protomer of *Te*CpcS-III in the crystallographic asymmetric unit.

Data collection and structure determination

A single-wavelength anomalous diffraction data set to 2.8 Å resolution was collected on a single crystal of the *Te*CpcS-III free enzyme at the X4A beamline of the National Synchrotron Light Source. The diffraction images were processed with the HKL package [45]. The data processing statistics are summarized in Table 2.

The selenium sites were located with the program SnB [46]. SOLVE/RESOLVE [47] was used for phase calculation, phase improvement and automated model building, but only about 8% of the residues were placed. The complete atomic models were built with the program XtalView [48], followed by structure refinement using the program CNS (Crystallography and NMR System) [49]. The refinement statistics are summarized in Table 2.

Structure analysis

Visualization and comparison of protein structures and manual docking of ligand molecule were performed using PyMol [50]. XtalView [48] and CNS [49] were used for the protein docking exercises and minimization of the steric clashes between the ligand and the protein. The overlaid structures of *Te*CpcS and UnaG bound to bilirubin (BR) reveal that BR resides nicely inside the barrel of *Te*CpcS. We also realized that the structure of UnaG provides insight into the conformation of two unstructured loops (residues 77–83 and 108–117) in the structure of *Te*CpcS. We therefore used the structure of UnaG (PDB id: 4I3B) as a template for modeling PCB and the two missing loops for *Te*CpcS. Subsequent to the manual building of the two missing loops and docking of the PCB in *Te*CpcS, the resulting model was subject to energy minimization using CNS.

Expression and purification of recombinant proteins for enzyme analyses

Plasmids were co-transformed into *E. coli* BL21 (DE3), and cells were grown on Luria-Bertani (LB) medium containing the appropriate antibiotics for selection as listed in Table 1 at the following concentrations: ampicillin, 100 μg ml^{-1}; chloramphenicol, 34 μg ml^{-1}; kanamycin, 50 μg ml^{-1}; and/or spectinomycin, 100 μg ml^{-1}. Isolated colonies were used to inoculate cultures (50-ml of LB medium amended with necessary antibiotics combinations at the aforementioned concentrations), which were grown at 37 °C with shaking at 225 rpm. This starter culture was used to inoculate 1 L of LB medium containing antibiotics for incubation until the culture reached an optical density $OD_{600\ nm} = 0.6$ (approximately 4 h). The temperature was lowered to 30 °C before induction of T7 RNA polymerase by addition of 1 mM IPTG, and the culture was incubated with shaking for an additional 3 h. Cells were collected by centrifugation at 10,000 × *g* for 10 min and stored at −20 °C until required. For PEB and PΦB production using pPebS and pHy2 expression plasmids, respectively, cultures were induced at 18 °C with 1 mM IPTG and incubated for 16 h prior to harvesting cells by centrifugation.

Frozen cell pellets were thawed and resuspended in a buffer of 50 mM Tris-HCl, 150 mM NaCl, pH 8.0 at 2.5 ml buffer per g wet-weight of cells, and the resulting cell suspension was lysed by three passages through a chilled French pressure cell at 138 MPa. Inclusion bodies, cell debris, and unbroken cells were removed by centrifugation for 20 min at 13,000 × *g*. For purification of hexa-histidine tagged proteins, the supernatant containing soluble proteins was applied to a Ni-NTA superflow affinity column (Qiagen, Inc., Chatsworth, CA) as previously described [18]. After elution of the desired protein with imidazole, the purified protein fraction was dialyzed overnight in imidazole-free suspension buffer containing 10 mM 2-mercaptoethanol.

Fluorescence emission and absorbance spectra

Absorbance spectra were acquired using a lambda 35, dual-beam UV/Vis spectrophotometer, and fluorescence emission and excitation spectra were recorded with a Perkin Elmer LS55 fluorescence spectrophotometer (Waltham, MA) as described [13]. The excitation wavelength was 490 nm for recombinant proteins carrying PEB chromophores and 590 nm for those carrying PCB or PΦB chromophores.

Protein and bilin analysis

Polypeptides were resolved by polyacrylamide gel electrophoresis (PAGE; 15% w/v) in the presence of sodium dodecyl sulfate (SDS), and polypeptides were visualized by staining with Coomassie blue as described [18]. To detect proteins containing bound chromophores (PEB, PCB or PΦB), gels were soaked in 10 mM $ZnSO_4$ [51]. The resulting enhanced fluorescence produced by chelation of the bilin by Zn was visualized using an FX imaging system (BioRad, Hercules, CA). Excitation at 532 nm was used to detect all three bilins, but excitation at 635 nm was used to detect PCB and PΦB.

Size exclusion chromatography (SEC)

SEC was performed using the protocol described earlier [16]. The molecular mass of native *Te*CpcS was calculated from a standard curve derived from a set of molecular mass standards (Bio-Rad).

Mass Spectrometric analysis

Purified *Te*CpcS was digested with trypsin, and the resulting tryptic peptides were analyzed using MALDI mass spectrometer (MS) and tandem MALDI MS/MS on a 4800 MALDI-ToF/ToF (AB Sciex, Concord, Ontario) following the procedures described previously [28].

RESULTS

Structural analysis of TeCpcS

The X-ray crystal structure of CpcS-III (hereafter *Te*CpcS) from *Thermosynechococcus elongatus* strain BP-1 (tll1699) was determined to 2.8-Å resolution; the coordinates and the structure factors were released by Protein Data Bank (PDB id: 3BDR) [38]. The structure reveals that *Te*CpcS contains two α-helices (Fig. 1) and one 10-stranded, antiparallel β-sheet that adopts a beta-barrel fold, which is capped by α1 at one side, and is widely open at the other side, presumably for the proper delivery of the substrate to the target phycobiliprotein.

*Te*CpcS is found as a homodimer in the crystal, and it belongs to the lipocalin structural family (Fig. 1). Various lipocalins exhibit different oligomeric states and occur as monomers, homodimers, heterodimers, or tetramers. Other members of this protein family also bind ligands, including fatty acids, retinol, carotenoids, pheromones, prostaglandins, biliverdin and bilirubin [52–62]. In the crystal structure, a phosphate ion forms a hydrogen-bond with the side chain of invariant Arg-151 located near the bottom of the funnel-like cavity. This interaction suggests that Arg151 is possibly involved in the substrate recognition.

The structure of *Te*CpcS was originally undertaken as part of a large-scale effort of the NIH Protein Structure Initiative to provide structural representatives from large domain families for which no structures were yet available. The aim of this program was to provide structural templates which could be leveraged by large-scale homology modeling [37, 63]. *Te*CpcS is the first structure of a protein from PFAM family PF09367, which includes more than 209 proteins and protein domains. This number is conservative because PFAM only considers select proteomes. A Hidden Markov Model (HMM) analysis of the 21.4 M distinct UniProt sequence using the PF09367 HMM provided by PFAM shows that there are 288 distinct sequences that have this domain signature. Based on Uniprot Version 2013_04 (w/~ 21.4 M unique sequences), this structure has a total modeling-level coverage (E val $1e^{-10}$) of 317 proteins and a novel modeling-level coverage [defined by [64]] of 317 proteins. Hence, the *Te*CpcS structure provides an important template for structural studies of this biologically important domain family, both by homology modeling and X-ray crystallography using molecular replacement.

According to an analysis using DALI [65], *Te*CpcS now has numerous structural homologs in the PDB, the top 55 of which have Z-scores ranging from 12.4 to 10.0. Most of these proteins are fatty acid binding proteins that are functionally unrelated to *Te*CpcS. In contrast, the closest structural homolog that is functionally related to *Te*CpcS is the fluorescent protein UnaG [62] (PDB id: 4I3B, Z-score 10.1, RMSD 2.8 over 109 residues). UnaG is a fluorescent protein recently discovered in freshwater eel (unagi) muscle that becomes fluorescent via noncovalent but tight binding to unconjugated bilirubin. This protein is hypothesized to aid in muscle metabolism of juvenile eels which must travel long distances and may benefit from the known antioxidant effects of bilirubin.[62] Interestingly, when the structure of UnaG bound to a bilirubin is overlaid onto the structure of *Te*CpcS, the substrate can readily be docked into the cavity of *Te*CpcS. This alignment only produced negligible clashes with the side chains of a few residues that are inside the funnel. This remarkable fit is the consequence of both proteins having similar β-barrel topologies, although they only share 11% sequence identity. More importantly, the structure of UnaG provides insight into the conformation of two unstructured loops (residues 77–83 and 108–117) in the structure of *Te*CpcS. The structure of UnaG suggests that these two loops become ordered only in presence of the substrate, because they interact with the substrate while partially capping the funnel. We therefore used the structure of UnaG (PDB id: 4I3B) as a template for modeling PCB and the two missing loops for *Te*CpcS (Fig. 2A).

Using the structure of the UnaG and its interaction with bilirubin as a guide, manual docking of PCB to *Te*CpcS was performed (Fig. 2A) by XtalView [48]. We performed the docking exercises by taking advantage of two important factors: 1) most bilins are negatively charged at neutral pH and 2) a significant number of residues lying inside the barrel are highly conserved and predominantly positively charged (see Fig. 3). Once a preliminary model was achieved, a refinement step using CNS was performed [49] to minimize possible polar and non-polar clashes between the protein and the substrate. The manual docking of UnaG in complex with bilirubin (BR) (PDB id: 4I3B) onto the structure of *Te*CpcS reveals that BR fits well in the cavity of *Te*CpcS (Fig. 2B). More importantly, the overlaid structures show that several residues that recognize the substrate BR in UnaG have close counterparts in *Te*CpcS, namely Arg112 and Arg132 in UnaG are closely aligned to Arg22 and Arg153 in *Te*CpcS, respectively. Additionally, there are several hydrophobic residues that interact with BR in UnaG, which have representatives in *Te*CpcS. We therefore took advantage of these striking similarities between the two proteins and modeled PCB based on what we observed in the overlaid structures.

As seen in Fig. 2A, the D ring of the model substrate is buried at the bottom of the funnel, where its carbonyl group is hydrogen bonded to the side chain of invariant Arg151. The position of this carbonyl group is similar to that of the phosphate ion in the crystal structure of the substrate-free *Te*CpcS. While the hydrophobic moieties of PCB interact with side chains of several conserved hydrophobic amino acids, namely Phe10, Phe11, Ile41, Leu90, Leu130, the propionate group on the C ring of PCB forms hydrogen bonds with invariant Trp73 and Ser155. Interestingly, the modeling of two missing loops suggests that the invariant Trp79 possibly caps the A ring and Tyr110 forms π-π interactions with the A ring. The structure of the PCB model docked into the cavity of *Te*CpcS is shown with the surface charge representation in Fig. 2C. The vinyl group on ring A of PCB is exposed and projects into the solvent, and there is a charge distribution difference along the surface of *Te*CpcS which may help to explain the substrate preference for CpcB, ApcA, ApcB, ApcD, and ApcF while excluding CpcA. Although these substrates are very similar at the structural level, one difference in charge near the chromophore attachment site may allow one subunit to bind while precluding CpcA from binding to *Te*CpcS. In all of the substrates for *Te*CpcS, the PCB binding motif CxRDx is followed by a negatively charged amino acid, either E or D, whereas in CpcA the CxRDx motif is followed by an uncharged G residue. A positively charged region containing K or R residues precedes the CxRDx motif in all phycobiliproteins subunits.

*Te*CpcS is 182 amino acids in length and is most related in sequence to the CpcS-III lyase from *Nostoc* sp. PCC 7120 (64.1% identity, 74% similarity) and to CpcS-I (45.5% identity, 60.6% similarity) and CpcU (28.9% identity; 44.7% similarity) from *Synechococcus* sp. PCC 7002; it is also similar in sequence to other S-type bilin lyases which attach PEB to phycoerythrins from *Prochlorococcus* spp. and *Fremyella diplosiphon* (see Supplementary Table 1). CpcS-I and CpcU form a heterodimeric bilin lyase that can ligate PCB to Cys82 (or equivalent position) on four types of AP subunits (ApcA, ApcB, ApcD, and ApcF) and CpcB [13, 16]. When *Te*CpcS was aligned with the sequences from *Synechococcus* sp. PCC 7002 CpcS-I and CpcU, the sequences were highly conserved across the entire sequence until the extreme C-terminus (Fig. 3) Although the gene encoding *Te*CpcS is the only open reading frame that shows similarity to a CpcS-type bilin lyase in the *T. elongatus* genome, its functionality as a bilin lyase had not been demonstrated.

Analysis of lyase activity of TeCpcS with CpcB

The bilin lyase activity of *Te*CpcS towards various apo-PBP subunits was tested using a previously developed heterologous coexpression system in *E. coli* [13]. Recombinant histidine-tagged proteins were purified using Ni-NTA affinity chromatography. The purified PBPs obtained from the *E. coli* cells coproducing CpcB and CpcA subunits (CpcBA), *Te*CpcS and enzymes to synthesize PCB from heme were analyzed by absorbance and fluorescence spectroscopy. Fig. 4A shows the absorbance spectrum (solid blue line, absorbance maximum at 620 nm) and fluorescence emission spectrum (dotted blue line, emission maximum at 644 nm) of PCB ligated to CpcB by *Te*CpcS (see Table 3). The black solid and dotted lines correspond to absorbance and fluorescence spectra, respectively, for purified CpcB/CpcA heterodimers obtained from cells producing PCB but in the absence of *Te*CpcS. The activity of the *Te*CpcS lyase was very similar to that of the CpcS-I/CpcU lyase, and red solid and dotted lines correspond to absorbance and fluorescence spectra, respectively, of CpcB chromophorylated at Cys82 by CpcS-I/CpcU (Fig. 4A; Table 3). The CpcB protein contains two PCB attachment sites, but the spectral properties of each ligated PCB are distinct: PCB ligated at Cys153 has maximal absorbance at 592 nm and maximal fluorescence emission at 624 nm [13, 18], whereas PCB ligated at Cys82 has an absorbance maximum at 621 nm and maximal fluorescence emission at 644 nm [13, 16] (see Table 3). Therefore, we conclude that *Te*CpcS is a PCB lyase with the same specificity as CpcS-I/CpcU.

To exclude the possibility that *Te*CpcS was attaching PCB to Cys84 of CpcA, the purified, PCB-containing CpcB/CpcA complex was resolved by SDS-PAGE (Fig. 4B). Covalent bilin addition to proteins was verified by fluorescence emission of the same gel after incubation with Zn sulfate as shown in Fig. 4C. The strong Zn-enhanced fluorescence emission observed for CpcB in lane 3 indicates that *Te*CpcS ligated PCB to CpcB but not to CpcA. A similar result was obtained when the lyase was CpcS-I/CpcU (Fig. 4C, lane 2). In the absence of any lyase, no Zn-enhanced fluorescence emission was detected (Fig. 4C, lane 1). There was no bilin addition to CpcA in any of the lanes shown in Fig. 4C. From these experiments, it can be concluded that *Te*CpcS acts as an S-type lyase that ligates PCB to CpcB at Cys82.

From the standpoint of potential biotechnological applications, it was also interesting to test the *Te*CpcS lyase activity with the non-cognate bilin substrates, PEB and PΦB. CpcB/CpcA and *Te*CpcS were produced in cells harboring pPebS or pHy2, which direct the production of PEB and PΦB from heme, respectively. As shown in Supplementary Fig. 1A, *Te* CpcS could attach PEB (solid red lines for absorbance, dotted red lines for fluorescence emission) and PΦB (solid green lines for absorbance and dotted green lines for fluorescence) to CpcB to produce highly fluorescent products with interesting spectral properties (see Table 3). These purified proteins were resolved by SDS-PAGE and stained with Coomassie Blue (Supplementary Fig. 1B). The bilin content of CpcA and CpcB was examined by Zn-enhanced fluorescence of the same gel with excitation at 532 nm (Supplementary Fig. 1C) or at 635 nm (Supplementary Fig. 1D), which detects PEB and PΦB, respectively. The results indicate that CpcB carries covalently bound PEB (lane 1) and PΦB (lane 2) when CpcB and CpcA are coproduced with *Te*CpcS in the presence of the appropriate bilin, but no bilin ligation was detected to CpcA in either case. These experiments established that *Te*CpcS is a versatile bilin lyase that is capable of attaching both cognate and non-cognate bilins to CpcB.

Analyzing activity of TeCpcS lyase on major allophycocyanin subunits ApcA/ApcB

The heterodimeric CpcS-I/CpcU lyase can chromophorylate four AP subunits at Cys81 [16, 19], so the *Te*CpcS was tested for its ability to chromophorylate ApcA and ApcB. ApcA and ApcB were coproduced with *Te*CpcS in cells that were also producing enzymes to synthesize PCB, PEB, and PΦB (see Table 1). The expressed proteins were purified using Ni-NTA affinity chromatography and analyzed by absorbance and fluorescence emission spectroscopy, and the data are shown in Supplementary Fig. 2. The solid lines show the absorbance spectra, and the dotted lines show the fluorescence emission spectra of the purified ApcA and ApcB produced in the presence of PCB (Supplementary Fig. 2A), PEB (Supplementary Fig. 2B), or PΦB (Supplementary Fig. 2C). The spectra in Supplementary Fig. 2A are consistent with the correct addition of PCB as established in published studies [16, 19]. No significant bilin ligation occurred in absence of the *Te*CpcS lyase (data not shown). The strong absorbance and fluorescence emission of purified ApcA/ApcB complexes coproduced with PEB (Supplementary Fig. 2B) and PΦB (Supplementary Fig. 2C) was strong evidence that these bilins were also correctly attached to Cys81 residues. The purified proteins were separated on SDS-PAGE and sequentially stained with Zn sulfate and Coomassie blue (Supplementary Fig. 2D), which verified that both ApcA and ApcB subunits were present in equal amounts. In addition, as shown in Supplementary Fig. 2E, both ApcA and ApcB subunits carried covalently bound, fluorescent PCB (lane 1), PEB (lane 2) or PΦB (lane 3). These data establish that *Te*CpcS can ligate PCB, PEB and PΦB to both ApcA and ApcB.

TeCpcS activity on AP α-like subunit ApcD

AP-B is an important terminal emitter of the PBS that transfers energy to Photosystem I and is composed of ApcB along with ApcD, a variant of the AP alpha subunit ($\alpha^{AP\text{-}B}$) [66–69]. CpcS-I/CpcU lyases ligate PCB to apo-ApcD when coproduced with PCB in *E. coli*. When ApcD is co-produced with ApcB, the solubility of ApcD is improved, and energy transfer from the bilin on ApcB to that on ApcD can be observed in the recombinant protein [13].

To test the bilin lyase activity of *Te*CpcS on ApcD, *Te*CpcS was co-expressed with ApcD/ApcB in the presence of one of the three bilins (PCB, PEB or PΦB). These resulting cells were intensely colored (data not shown), which suggested that covalent bilin ligation had occurred. The His-tagged PBPs were purified from cells using Ni-NTA affinity chromatography and characterized as described above using absorbance and fluorescence emission spectroscopy. The absorbance (solid lines) and fluorescence emission (dashed lines) in Supplementary Fig. 3A, B, and C show the results for ApcD/ApcB ligated with three different bilins: PCB (Supplementary Fig. 3A), PEB (Supplementary Fig. 3B), and PΦB (Supplementary Fig. 3C) (see Table 3). Interestingly, only when PCB was ligated to these proteins did the longer wavelength absorption band characteristic of native ApcD appear. To show that the chromophores were covalently bound to both ApcD and ApcB, the purified proteins were separated on SDS-PAGE and stained with Zn sulfate (Supplementary Fig. 3E, lanes 1–3) and subsequently with Coomassie blue (Supplementary Fig. 3D). Zn-enhanced fluorescence showed that bilin chromophores were covalently attached to both ApcB and ApcD. Therefore, *Te*CpcS can attach PCB, PEB, and PΦB to ApcD.

TeCpcS activity on ApcF

ApcF is a variant type of AP β subunit (also known as β^{18}) that partners with the amino-terminal domain of ApcE (the core membrane linker PBP subunit), the second terminal emitter of PBS. In *Synechococcus* sp. strain PCC 7002, the loss of ApcF affects energy transfer from PBS to photosystem II [67, 69–71]. A previous study showed that the heterodimeric CpcS-I/CpcU lyase can ligate PCB on ApcF [13]. *Te*CpcS was coproduced with ApcF together with enzymes to produce three bilins (PCB, PEB or PΦB and a combination of both PCB and PEB). The resulting cells were intensely colored (data not shown), which suggested that efficient ligation of chromophores to ApcF had occurred. The proteins were purified using Ni-NTA affinity chromatography and analyzed. In Supplementary Fig. 4A, the solid lines represent absorbance and dotted lines represent the fluorescence emission spectra of ApcF carrying PCB (blue) or PEB (red; Table 3). Next, a competition experiment with *Te*CpcS was performed in which cells co-expressed ApcF, *Te*CpcS and the genes required to make both PCB and PEB. Even though PCB is the cognate bilin in *T. elongatus*, *Te*CpcS was unable to discriminate between PCB vs. PEB, as the absorbance spectrum shown in purple shows that both bilins were attached to ApcF (see Supplementary Fig. 4B). Fluorescence emission from PEB (red dotted line) and PCB (blue dotted line) was observed as well, indicating that both bilins had been ligated to ApcF. *Te*CpcS was also capable of attaching PΦB to ApcF (Supplementary Fig. 4C). The purified proteins were separated on SDS-PAGE and stained with Zn sulfate to confirm bilin addition to ApcF (Supplementary Fig. 4E, lanes 1 through 4). *Te*CpcS was able to attach PCB, PEB, and PΦB to ApcF, even though ApcF naturally carries only PCB.

Intrinsic bilin binding of TeCpcS

The purified HT-*Te*CpcS obtained by producing the protein together with PCB or PEB was weakly fluorescent (Fig. 5A and 5B). However, the fluorescence intensity of *Te*CpcS was much lower than that obtained with PBP subunits. After electrophoresis in the presence of SDS, both *Te*CpcS-PCB and *Te*CpcS-PEB had readily detected Zn-enhanced fluorescence emission bands (Fig. 5D, lanes 2 and 3, respectively). To understand the nature of the interaction of the bilin to *Te*CpcS, the HT-*Te*CpcS-PCB and HT-*Te*CpcS-PEB adducts were analyzed by trypsin digestion and mass spectrometry. HT-*Te*CpcS bound both PEB and PCB at Cys2 and Cys169 (Supplementary Fig. 5). Based on the X-ray structure and modeling (Fig. 2A and Fig. 3), the bilin is covalently bound to *Te*CpcS within the cavity of the β-barrel at Cys169 (location labeled in Fig 2A) and at the N-terminal loop at Cys2 near the first helix. Neither of these Cys residues is conserved within the CpcS-I/U family (see Fig. 3), which suggests that covalent ligation to these Cys residues may occur adventitiously during the purification process. In the cyanobacterial cytoplasm, it is unlikely that a PCB molecule would have a long association with the *Te*CpcS enzyme. The function of *Te*CpcS is to ligate PCB to apo-phycobiliprotein substrates, and these substrate proteins are the most abundant proteins in cyanobacterial cells under most growth conditions. Therefore, it seems unlikely that the covalent addition is a normal occurrence in cyanobacteria or a normal part of the reaction cycle.

To test whether these two Cys residues were involved in catalysis, site-specific variants within the *Te*CpcS were produced (C2S, C169S, and C2S/C169S). These HT-*Te*CpcS variants were produced in *E. coli* by co-expressing them with pPcyA, purified, and analyzed by absorbance and fluorescence emission spectroscopy followed by SDS-PAGE and Zn-enhanced bilin fluorescence. Wild-type *Te*CpcS and all of the variants retained the ability to bind PCB as judged by their absorbance and fluorescence spectra (Supplementary Fig. 6A; absorbance maximum at 605 nm and fluorescence emission maximum at 633 nm). However, when these protein/bilin complexes were separated by SDS-PAGE (Supplementary Fig. 6C) and assayed for their covalent binding of PCB by Zn-enhanced fluorescence of *Te*CpcS, the C2S and the C2S/C169S variants had no detectable PCB fluorescence associated with the *Te*CpcS protein (lanes 1–4 in Supplementary Fig. 6D). This indicates that covalent addition of PCB to *Te*CpcS mainly occurs through Cys2.

In order to assay whether these site-specific variants retained their PCB ligation activity, the genes encoding the *Te*CpcS variants were co-expressed with genes encoded by pPcyA and pCpcBA. In this experiment both *Te*CpcS and CpcB were His-tagged and copurified. The absorbance and fluorescence emission properties of the purified proteins from these co-expressions are consistent with those of HT-CpcB with PCB covalently attached at Cys82 (see Supplementary Fig. 6B) with an absorbance maxima at 621 nm and fluorescence maxima at 640 nm[16]. These samples containing HT-CpcB were also diluted approximately 4-fold in comparison to those shown in Supplementary Fig. 6A. Although the masses of HT-*Te*CpcS and HT-CpcB are very similar and they co-migrate in SDS-PAGE (see lanes 5–8 in Supplementary Fig. 6C and 6D), the large amount of bilin fluorescence seen in the samples containing all of the variants indicates that they are capable of PCB ligation to HT-CpcB. Therefore, we conclude that neither Cys2 nor Cys169 is required for the activity of *Te*CpcS.

In the *Te*CpcS model with bound PCB, PCB is suggested to have hydrophobic interactions as well as hydrogen bonding with numerous residues. According to the model, both Arg151 and Ser155 interact with PCB, forming hydrogen bonds to the D and C rings, respectively (see Fig. 2A). Therefore, site-specific variants R151G and S155G were generated within *Te*CpcS and tested for their ability to bind PCB. These HT-*Te*CpcS variants were produced in *E. coli* by co-expressing them with pPcyA, purified, and analyzed by absorbance and fluorescence emission spectroscopy, which was followed by SDS-PAGE and Zn-enhanced bilin fluorescence. Both the wild-type *Te*CpcS and the S155G variant retained the ability to bind PCB as seen in Supplementary Figs. 7A, B and C. The S155G variant has less PCB attached to *Te*CpcS than the wild-type, suggesting the strength of the interaction with PCB may be weaker (compare lanes 1 and 3 in Supplementary Fig. 7C). However, the R151G variant had no detectable PCB fluorescence (Supplementary Fig. 7A, orange spectra), and covalent PCB attachment to *Te*CpcS was not detected by Zn-enhanced fluorescence (Supplementary Fig. 7C, lane 2). Therefore we conclude that both residues appear to be involved in the binding of PCB to *Te*CpcS, consistent with the model in Fig. 2A. However, when these variants were co-expressed with pPcyA and pCpcBA to test for PCB ligation activity, both variants were capable of ligating PCB to HT-CpcB (Supplementary Figs. 7D, E, and F). Therefore, even though both *Te*CpcS variants S155G and R151G appear to contain less PCB than wild-type after SDS-PAGE, these variants still bind PCB sufficiently well to allow attachment to CpcB to occur in *E. coli*.

Investigation of the native molecular weight of the TeCpcS protein

In order to determine the oligomeric status of *Te*CpcS in its native state, purified recombinant HT-*Te*CpcS was subjected to size exclusion chromatography (Fig. 6). The protein complex had a retention time of 26.9 min, and the molecular weight of this complex was calculated to be 45,600 (Fig. 6, inset graph). The calculated molecular mass of the HT-*Te*CpcS polypeptide is 21.9 kDa. These data suggest that the native protein is stable and active as a homodimer, which is consistent with the observation that *Te*CpcS crystallizes as a dimer (Fig. 1).

DISCUSSION

We describe here the first structure for a bilin lyase of the CpcS/CpcU family and demonstrate its functionality and substrate capabilities in a recombinant, heterologous expression system. *Te*CpcS is a member of a diverse family of proteins known as the lipocalins. All members of the lipocalin family are composed of similar beta-barrels, and they bind a diverse array of ligands, most of which are hydrophobic. Two members of the lipocalin family of proteins bind bilins, UnaG and the bilin-binding protein of insects that binds biliverdin IXγ. Interestingly, the motivation for solving the structure of the bilin-binding protein of insects in 1987 was to compare it to the structure of PBPs, because both of these proteins associated with bilins. However, they were not structurally related in any way: PBPs are most similar to hemoglobin and myoglobin, while the insect bilin-binding protein was most similar to the retinol-binding protein. Very recently, the structure of UnaG, the first fluorescent protein found in vertebrates, was solved [62]. This protein is most similar to fatty acid-binding proteins and is fluorescent because it non-covalently binds bilirubin. UnaG is only fluorescent when bilirubin is bound and is not fluorescent when the similar bilin, biliverdin IXα, is bound [62]. UnaG and *Te*CpcS share structural similarities and both proteins bind bilins. Our modeling suggests that the association of PCB with *Te*CpcS (Fig. 2) is similar to that shown for UnaG and bilirubin. The structure of UnaG allowed us to predict the conformation of two unstructured loops (residues 77–83 and 108–117) that were missing in the electron density map for *Te*CpcS. These two loops may become ordered only in presence of the bilin, because they interact with the bilin while partially capping the funnel of the β-barrel. The location of these loops near the bilin and near the wide-end of the "funnel" also suggests that one or both of these loops may also be involved in phycobiliprotein substrate docking. An examination of the alignment in Fig. 3 reveals some substantial differences in the region of 108–117 between the CpcS-type lyases and the CpeS lyase. The CpcS-type lyases in the alignment ligate PCB to allophycocyanin and phycocyanin β-subunits, whereas the CpeS lyase ligates PEB to phycoerythrin β-subunits (CpeB). It is also possible that the structures of these loops in *Te*CpcS are very different from those in UnaG. The function of UnaG in eel is not currently known, but it does not appear to have a catalytic role in bilin ligation as we show here for *Te*CpcS.

The CpcE/CpcF bilin lyase family members are not phylogenetically related to the CpcS/CpcU family [17, 21, 32, 40]. All CpcE/CpcF-type lyases contain 5–6 HEAT-repeat motifs [72, 73], which are also found in diverse proteins from various eukaryotic organisms and are thought to facilitate protein-protein interactions [74, 75]. No structures have been solved for these CpcE/CpcF lyase proteins, but Phyre^2 analyses [76] of a fusion protein MpeZ from this group suggests that their structures are primarily alpha-helical [31].

Phyre^2 analyses of the proteins within the CpcT lyase family suggest that they are distantly related to the CpcS/CpcU proteins [40]. It was hypothesized that lyases of the CpcT-type were necessary in order to form the *S*-stereoisomer at C3^1 of the bilin located at Cys153 on CpcB, because bilins attached by the CpcS/CpcU-type lyases have *R*-configuration at C3[118]. Presumably, if members of the CpcT family are also members of the same structural family as *Te*CpcS, then the bilin would be held in a different conformation within the beta-barrel to allow the *S*-epimer to form. Evolution of the CpcE/CpcF family of lyases may have occurred later, as suggested by Biswas *et al*. [28], in order to create a set of lyases that was highly specific for the alpha subunits only, and these lyases later evolved an isomerase activity to allow a more diverse array of bilins to be produced with these antenna proteins.

*Te*CpcS is a member of the CpcS/CpcU family of bilin lyases, and we show here that in *E. coli*, it can ligate a diverse array of bilins to its PBP substrates CpcB and allophycocyanin subunits ApcA, ApcB, ApcD, and ApcF at the equivalent position to Cys82. Members of the CpcS/CpcU lyase family all seem to bind bilins rapidly and tightly, and then catalyze a slow thioether bond formation to the PBPs. This appears to be very important to prevent spontaneous but incorrectly ligated products from forming [35, 77–79]. These lyases play a chaperone-like role in which they quickly and specifically bind the bilin and deliver it to the PBP substrate in an appropriate conformation for ligation [77, 80, 81].

The high reactivity of bilins towards cysteines precluded initial attempts to obtain co-crystals of PCB with *Te*CpcS probably due to the mixed population of covalent adducts that was formed. This problem could possibly be overcome by mutating the cysteine residues or by using a bilin substrate analog without vinyl groups on the A and D rings. As we demonstrate here, bilins were covalently bound to Cys2 and Cys169 on *Te*CpcS (Supplementary Fig. 5), but neither of these Cys is highly conserved as seen in the alignment in Fig. 3, and both are very far from the target vinyl group. We also show here that site specific variants C2S, C169S and C2S/C169S variants were capable of PCB ligation to HT-CpcB, indicating that neither Cys is involved in the catalytic mechanism.

In the structure of *Te*CpcS modeled with PCB, the target vinyl group is exposed in a shallow cavity on the surface of the protein (see Fig. 2C). *Te*CpcS holds it in a precise geometry ready for catalysis, but the active site would need to be completed by the binding of the phycobiliprotein subunit that brings its target Cys residue into the groove on the surface of *Te*CpcS to allow nucleophilic attack by the cysteinyl sulfur on the vinyl group on ring A of PCB. This would mean the enzyme active site would actually be divided between *Te*CpcS and the Cys on the target phycobiliprotein subunit, and this could explain why chromophore ligation is much slower than bilin binding.

A histidine was suggested as a candidate amino acid involved in the binding/ligation reaction mechanism in phytochromes [82]. Tu *et al*. suggested that His residues were the most likely candidate for the strong association of PCB with the lyase, because they were unable to detect covalent addition with Cys after trypsin digestion in their mass spectrometry experiments [83]. The results of Kupka *et al*. suggested when these histidines in CpcS-III from *Nostoc* sp. PCC 7120 were mutated, the conformation and or protonation state of the chromophore was affected [77]. Cytochrome *c* biogenesis requires the heme chaperone CcmE which was shown to bind heme covalently through an association with histidine residues [84]. However, in the modeling with PCB shown in Fig. 2, no histidines occur in locations to allow a strong interaction with the bilin. Studies performed on the *Nostoc* sp. PCC 7120 CpcS-III protein showed that two arginines (Arg18 and Arg149, numbering from 7120 CpcS-III) were essential for lyase activity. In our model with *Te*CpcS and PCB, Arg151 (equivalent to Arg149 in the *Nostoc* sp. protein; see Fig. 3) was modeled to hydrogen bond to the carbonyl group of ring D [77]. When Arg151 was mutated to glycine, *Te*CpcS did not retain PCB after purification with the enzyme, unlike the wild-type protein, confirming its importance in the stable association between *Te*CpcS and PCB. However, this variant retained the ability to ligate PCB to HT-CpcB, indicating that PCB can still bind to this variant of *Te*CpcS.

Like some other lyases, under the heterologous conditions employed here (i.e., in *E. coli*), *Te*CpcS displayed an ability to attach both cognate as well as non-cognate bilins to various PBP subunits [23, 24]. *In vivo*, it was also demonstrated that co-expression of the genes required to make PEB in *Synechococcus* sp. PCC 7002 resulted in the attachment of the non-cognate PEB to CpcA (by the CpcE/CpcF lyase) while expression of the genes required to produce PΦB resulted in many phycobiliproteins containing the PΦB chromophore [85]. *T. elongatus* is a cyanobacterial species that only contains PCB, so its lyases do not need to have a strong discrimination between the isomeric bilins PEB and PCB. Other lyases, like CpeS from *Fremyella diplosiphon*, are very specific with respect to their bilin substrate, specifically attaching PEB but not PCB at Cys80 on CpeB [28], and this ability to discriminate presumably evolved to insure that the appropriate bilin is attached to the correct protein so that energy transfer within PBS is efficient and unidirectional. *Te*CpcS was shown to ligate three different bilins (its cognate bilin, PCB, as well as PEB and PΦB) to Cys82 of five different apo-PBP substrates (CpcB, ApcA, ApcB, ApcD, and ApcF). This capability allows the production of a diverse array of fluorescent, natural and unnatural phycobiliproteins for protein tagging and bioimaging purposes. Thus, the findings presented here illustrate the utility of this thermostable enzyme for producing fluorescent proteins for future biotechnological applications.

CONCLUSION

*Te*CpcS is a bilin lyase of the CpcS family and has a beta-barrel structure that is similar to those of diverse members of the lipocalin family. *Te*CpcS forms homodimers in solution and can ligate the cognate bilin, PCB, as well as two non-cognate bilins PEB and PΦB) to Cys82 of five different PBP substrates, including CpcB but notably not CpcA. The crystal structure and an energy-minimized model with bound substrate provide the structural basis for future studies to understand how specificity is generated within this class of enzymes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding Source Statement:

This work was funded by grants from the Protein Structure Initiative of the National Institutes of Health to G. T. M. and J. F. H. (U54 GM094597) and from the National Science Foundation to W. M. S. (MCB-0843664 and MCB-1244339), D. A. B. (MCB-0519743 and MCB-1021725), and R.B.C. (CHE-1058764).

We thank Randy Abramowitz and John Schwanof for setting up the X4A beamline; Sergey M. Vorobiev for assistance in crystal screening. The W. M. Keck Foundation provided support for equipment utilized for this study and located in the Keck Conservation and Molecular Genetics lab at the University of New Orleans.

Abbreviations

AP	allophycocyanin
DTT	dithiothreitol
IPTG	isopropyl-β-D-thiogalactopyranoside
LB	Luria-Bertani medium
MALDI	matrix-assisted laser desorption ionization
MS	mass spectrometry
Ni-NTA	nickel nitrolotriacetic acid
PAGE	polyacrylamide gel electrophoresis
PBP	phycobiliprotein(s)
PBS	phycobilisome(s)
PC	phycocyanin
PCB	phycocyanobilin
PCR	polymerase chain reaction
PEB	phycoerythrobilin
PΦB	phytochromobilin
SDS	sodium dodecylsulfate
SEC	size exclusion chromatography
Te	*Thermosynechococcus elongatus*
ToF	time of flight

References

1. Glazer AN. Light guides: directional energy transfer in a photosynthetic antenna. J Biol Chem. 1989; 264:1–4.
2. Willows RD, Mayer SM, Foulk MS, DeLong A, Hanson K, Chory J, Beale SI. Phytobilin biosynthesis: the *Synechocystis* sp. PCC 6803 heme oxygenase-encoding *ho1* gene complements a phytochrome-deficient *Arabidopsis thaliana* hy1 mutant. Plant Mol Biol. 2000; 43:113–120.
3. Rhie G, Beale SI. Phycobilin Biosynthesis: Reductant Requirements and Product Identification for Heme Oxygenase from *Cyanidium caldarium*. Arch Biochem Biophys. 1995; 320:182–194. [PubMed: 7793979]
4. Cornejo J, Willows RD, Beale SI. Phytobilin biosynthesis: cloning and expression of a gene encoding soluble ferredoxin-dependent heme oxygenase from *Synechocystis* sp. PCC 6803. Plant J. 1998; 15:99–107. [PubMed: 9744099]
5. Rhie G, Beale SI. Biosynthesis of Phycobilins: Ferredoxin-Supported NADPH-Independent Heme Oxygenase and Phycobilin-Forming Activities from *Cyanidium caldarium*. J Biol Chem. 1992; 267:16088–16093. [PubMed: 1644795]
6. Kohchi T, Mukougawa K, Frankenberg N, Masuda M, Yakota A, Lagarias JC. The *Arabidopsis* HY2 gene encodes phytochromobilin synthase, a ferredoxin-dependent biliverdin reductase. Plant Cell. 2001; 13:425–436. [PubMed: 11226195]
7. Frankenberg N, Mukougawa KK, Lagarias JC. Functional genomic analysis of the HY2 family of ferredoxin-dependent bilin reductases from oxygenic photosynthetic organisms. Plant Cell. 2001; 13:965–978. [PubMed: 11283349]
8. Rockwell NC, Su YS, Lagarias JC. Phytochrome structure and signaling mechanisms. Annual Review of Plant Biology. 2006; 57:837–858.
9. Schirmer T, Huber R, Schneider M, Bode W, Miller M, Hackert ML. Crystal structure analysis and refinement at 2.5 Å of hexameric C-phycocyanin from the cyanobacterium *Agmenellum quadruplicatum*: the molecular model and its implications from light-harvesting. J Mol Biol. 1986; 188:651–676. [PubMed: 3090271]
10. Glazer AN, Fang S. Chromophore content of blue-green algal phycobiliproteins. J Biol Chem. 1973; 248:659–662. [PubMed: 4630849]
11. Capuano V, Braux AS, Tandeau de Marsac N, Houmard J. The “anchor polypeptide” of cyanobacterial phycobilisomes. Molecular characterization of the *Synechococcus* sp PCC 6301 *apcE* gene. J Biol Chem. 1991; 266:7239–7247. [PubMed: 1901865]
12. Zhao KH, Ping S, Bohm S, Bo S, Ming Z, Bubenzer C, Scheer H. Reconstitution of phycobilisome core-membrane linker, L-CM, by autocatalytic chromophore binding to ApcE. Biochim Biophys Acta-Bioenerg. 2005; 1706:81–87.
13. Biswas A, Vasquez YM, Dragomani TM, Kronfel ML, Williams SR, Alvey RM, Bryant DA, Schluchter WM. Biosynthesis of cyanobacterial phycobiliproteins in *Escherichia coli*: Chromophorylation efficiency and specificity of all bilin lyases from *Synechococcus* sp strain PCC 7002. Appl Environ Microbiol. 2010; 76:2729–2739. [PubMed: 20228104]
14. Zhao KH, Deng MG, Zheng M, Zhou M, Parbel A, Storf M, Meyer M, Strohmann B, Scheer H. Novel activity of a phycobiliprotein lyase: both the attachment of phycocyanobilin and the isomerization to phycobiliviolin are catalyzed by the proteins PecE and PecF encoded by the phycoerythrocyanin operon. FEBS Lett. 2000; 469:9–13. [PubMed: 10708746]
15. Fairchild CD, Zhao J, Zhou J, Colson SE, Bryant DA, Glazer AN. Phycocyanin α subunit phycocyanobilin lyase. Proc Natl Acad Sci USA. 1992; 89:7017–7021. [PubMed: 1495995]
16. Saunée NA, Williams SR, Bryant DA, Schluchter WM. Biogenesis of phycobiliproteins. II. CpcS-I and CpcU comprise the heterodimeric bilin lyase that attaches phycocyanobilin to Cys-82 of beta - phycocyanin and Cys-81 of allophycocyanin subunits in *Synechococcus* sp. PCC 7002. J Biol Chem. 2008; 283:7513–7522. [PubMed: 18199753]
17. Scheer H, Zhao KH. Biliprotein maturation: the chromophore attachment. Mol Microbiol. 2008; 68:263–276. [PubMed: 18284595]
18. Shen G, Saunee NA, Williams SR, Gallo EF, Schluchter WM, Bryant DA. Identification and characterization of a new class of bilin lyase: the *cpcT* gene encodes a bilin lyase responsible for

attachment of phycocyanobilin to Cys-153 on the beta subunit of phycocyanin in *Synechococcus* sp. PCC 7002. J Biol Chem. 2006; 281:17768–17778. [PubMed: 16644722]

19. Zhao KH, Su P, Tu J, Wang X, Liu H, Ploscher M, Eichacker L, Yang B, Zhou M, Scheer H. Phycobilin:cystein-84 biliprotein lyase, a near-universal lyase for cysteine-84-binding sites in cyanobacterial phycobiliproteins. Proc Natl Acad Sci USA. 2007; 104:14300–14305. [PubMed: 17726096]
20. Zhou J, Gasparich GE, Stirewalt VL, de Lorimier R, Bryant DA. The *cpcE* and *cpcF* genes of *Synechococcus* sp. PCC 7002: construction and phenotypic characterization of interposon mutants. J Biol Chem. 1992; 267:16138–16145. [PubMed: 1644801]
21. Schluchter, WM.; Shen, G.; Alvey, RM.; Biswas, A.; Saunée, NA.; Williams, SR.; Miller, CA.; Bryant, DA. Phycobiliprotein biosynthesis in cyanobacteria: Structure and function of enzymes involved in post-translational modification. In: Hallenbeck, PC., editor. Advances in Experimental Medicine & Biology. Springer; NY, NY: 2010.
22. Tang K, Zeng XL, Yang Y, Wang ZB, Wu XJ, Zhou M, Noy D, Scheer H, Zhao KH. A minimal phycobilisome: Fusion and chromophorylation of the truncated core-membrane linker and phycocyanin. Biochim Biophys Acta-Bioenerg. 2012; 1817:1030–1036.
23. Blot N, Wu XJ, Thomas JC, Zhang J, Garczarek L, Bohm S, Tu JM, Zhou M, Ploscher M, Eichacker L, Partensky F, Scheer H, Zhao KH. Phycourobilin in trichromatic phycocyanin from oceanic cyanobacteria is formed posttranslationally by a phycoerythrobilin lyase-isomerase. J Biol Chem. 2009; 284:9290–9298. [PubMed: 19182270]
24. Alvey RM, Biswas A, Schluchter WM, Bryant DA. Attachment of noncognate chromophores to CpcA of *Synechocystis* sp. PCC 6803 and *Synechococcus* sp PCC 7002 by heterologous expression in *Escherichia coli*. Biochemistry. 2011; 50:4890–4902. [PubMed: 21553904]
25. Bhalerao RP, Kind LK, Gustafsson P. Cloning of the *cpcE* and *cpcF* genes from *Synechococcus* sp. PCC 6301 and their inactivation in *Synechococcus* sp. PCC 7942. Plant Mol Biol. 1994; 26:313–326. [PubMed: 7524727]
26. Fairchild CD, Glazer AN. Oligomeric structure, enzyme kinetics, and substrate specificity of the phycocyanin alpha subunit phycocyanobilin lyase. J Biol Chem. 1994; 269:8686–8694. [PubMed: 8132596]
27. Swanson RV, Zhou J, Leary JA, Williams T, de Lorimier R, Bryant DA, Glazer AN. Characterization of phycocyanin produced by *cpcE* and *cpcF* mutants and identification of an intergenic suppressor of the defect in bilin attachment. J Biol Chem. 1992; 267:16146–16154. [PubMed: 1644802]
28. Biswas A, Boutaghou MN, Alvey RM, Kronfel CM, Cole RB, Bryant DA, Schluchter WM. Characterization of the activities of the CpeY, CpeZ, and CpeS bilin lyases in phycoerythrin biosynthesis in *Fremyella diplosiphon* strain UTEX 481. J Biol Chem. 2011; 286:35509–35521. [PubMed: 21865169]
29. Storf M, Parbel A, Meyer M, Strohmann B, Scheer H, Deng MG, Zheng M, Zhou M, Zhao KH. Chromophore attachment to biliproteins: specificity of PecE/PecF, a lyase-isomerase for the photoactive 3(1)-cys-alpha 84-phycoviolobilin chromophore of phycoerythrocyanin. Biochemistry. 2001; 40:12444–12456. [PubMed: 11591166]
30. Jung LJ, Chan CF, Glazer AN. Candidate genes for the phycoerythrocyanin α subunit lyase: biochemical analysis of *pecE* and *pecF* interposon mutants. J Biol Chem. 1995; 270:12877–12884. [PubMed: 7759546]
31. Shukla A, Biswas A, Blot N, Partensky F, Karty JA, Hammad LA, Garczarek L, Gutu A, Schluchter WM, Kehoe DM. Phycoerythrin-specific bilin lyase-isomerase controls blue-green chromatic acclimation in marine *Synechococcus*. Proceedings of the National Academy of Sciences. 2012; 109:20136–20141.
32. Shen, GZ.; Saunee, NA.; Gallo, E.; Begovic, Z.; Schluchter, WM.; Bryant, DA. Identification of novel phycobiliprotein lyases in cyanobacteria. In: Niederman, RA.; Blankenship, RE.; Frank, H.; Robert, B.; van Grondelle, R., editors. Photosynthesis 2004 Light-Harvesting Systems Workshop. Saint Adele; Quebec, Canada: 2004. p. 14-15.
33. Shen G, Schluchter WM, Bryant DA. Biogenesis of phycobiliproteins. I. *cpcS-I* and *cpcU* mutants of the cyanobacterium *Synechococcus* sp. PCC 7002 define a heterodimeric phycocaynobilin lyase

specific for beta -phycocyanin and allophycocyanin subunits. J Biol Chem. 2008; 283:7503–7512. [PubMed: 18199754]

34. Zhao KH, Zhang J, Tu JM, Boehm S, Ploescher M, Eichacker L, Bubenzer C, Scheer H, Wang X, Zhou M. Lyase activities of CpcS- and CpcT-like proteins from *Nostoc* PCC7120 and sequential reconstitution of binding sites of phycoerythrocyanin and phycocyanin beta-subunits. J Biol Chem. 2007; 282:34093–34103. [PubMed: 17895251]
35. Wiethaus J, Busch AWU, Kock K, Leichert LI, Herrmann C, Frankenberg-Dinkel N. CpeS is a lyase specific for attachment of 3Z-PEB to Cys82 of β- phycoerythrin from *Prochlorococcus marinus* MED4. J Biol Chem. 2010; 285:37561–37569. [PubMed: 20876568]
36. Zhao KH, Su P, Li J, Tu JM, Zhou M, Bubenzer C, Scheer H. Chromophore attachment to phycobiliprotein beta-subunits: phycocyanobilin:cystein-beta84 phycobiliprotein lyase activity of CpeS-like protein from *Anabaena* sp. PCC7120. J Biol Chem. 2006; 281:8573–8581. [PubMed: 16452471]
37. Montelione GT. The Protein Structure Initiative: achievements and visions for the future. F1000 Biol Reports. 2012; 4:7.
38. Kuzin AP, Su M, Seetharaman J, Forouhar F, Wang D, Janjua H, Cunningham K, Ma L-C, Xiao R, Liu J, Baran MC, Acton TB, Rost B, Montelione GT, Tong L, Hunt JF. Northeast Structural Genomics Consortium. Crystal structure of fatty acid-binding protein-like Ycf58 from *Thermosynechoccus elongatus* target Ter13*pdb ID: 3BDR*. Target Ter13. 200710.2210/pdb3bdr/pdb
39. Nakamura Y, Kaneko T, Sato S, Ikeuchi M, Katoh H, Sasamoto S, Watanabe A, Iriguchi M, Kawashima K, Kimura T, Kishida Y, Kiyokawa C, Kohara M, Matsumoto M, Matsuno A, Nakazaki N, Shimpo S, Sugimoto M, Takeuchi C, Yamada M, Tabata S. Complete genome structure of the thermophilic cyanobacterium *Thermosynechococcus elongatus* BP-1. DNA Res. 2002; 9:123–130. [PubMed: 12240834]
40. Bretaudeau A, Coste F, Humily F, Garczarek L, Le Corguillé G, Six C, Ratin M, Collin O, Schluchter WM, Partensky F. CyanoLyase: a database of phycobilin lyase sequences, motifs and functions. Nucleic Acids Research. 2013; 41:D396–D401. [PubMed: 23175607]
41. Xiao R, Anderson S, Aramini JM, Belote R, Buchwald W, Ciccosanti C, Conover K, Everett JK, Hamilton K, Huang YJ, Janjua H, Jiang M, Kornhaber GJ, Lee DY, Locke JY, Ma L-C, Maglaqui M, Mao L, Mitra S, Patel D, Rossi P, Sahdev S, Sharma S, Shastry R, Swapna GVT, Tong SN, Wang D, Wang H, Zhao L, Montelione GT, Acton TB. The high-throughput protein sample production platform of the Northeast Structural Genomics Consortium. J Struct Biol. 2010; 172:21–33. [PubMed: 20688167]
42. Acton TB, Xiao R, Anderson S, Aramini JM, Buchwald W, Ciccosanti C, Conover K, Everett JK, Hamilton K, Huang YJ, Janjua H, Kornhaber GJ, Lau J, Lee DY, Liu G, Maglaqui M, Ma L-C, Mao L, Patel D, Rossi P, Sahdev S, Sharma S, Shastry R, Swapna GVT, Tang Y, Tong SN, Wang D, Wang H, Zhao L, Montelione GT. Preparation of protein samples for NMR structure, function, and small molecule screening studies. Methods Enzymol. 2011; 493:21–60. [PubMed: 21371586]
43. Jansson M, Li YC, Jendeberg L, Anderson S, Montelione GT, Nilsson B. High-level production of uniformly ^{15}N- and ^{13}C-enriched fusion proteins in *Escherichia coli*. J Biomol NMR. 1996; 7:131–141. [PubMed: 8616269]
44. Doublié S, Kapp U, Åerg A, Brown K, Strub K, Cusack S. Crystallization and preliminary X-ray analysis of the 9 kDa protein of the mouse signal recognition particle and the selenomethionyl-SRP9. FEBS Lett. 1996; 384:219–221. [PubMed: 8617357]
45. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. 1997; 276:307–326.
46. Weeks CM, Miller R. The design and implementation of SnB v2.0. J Appl Cryst. 1999; 32:120–124.
47. Terwilliger TC. Solve and resolve: automated structure solution and density modification. Methods Enzymol. 2003; 374:22–37. [PubMed: 14696367]
48. McRee DE. XtalView/Xfit-A versatile program for manipulating atomic coordinates and electron density. Journal of Structural Biology. 1999; 125:156–165. [PubMed: 10222271]

49. Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Cryst. 1998; D54:905–921.
50. DeLano, WL. The Pymol manual. DeLano Scientific; San Carlos, CA: 2002.
51. Berkelman TR, Lagarias JC. Visualization of bilin-linked peptides and proteins in polyacrylamide gels. Anal Biochem. 1986; 156:194–201. [PubMed: 3526971]
52. Bishop RE. The bacterial lipocalins. Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology. 2000; 1482:73–83.
53. Charron JBF, Ouellet F, Pelletier M, Danyluk J, Chauve C, Sarhan F. Identification, expression, and evolutionary analyses of plant lipocalins. Plant Physiol. 2005; 139:2017–2028. [PubMed: 16306142]
54. Grzyb J, Latowski D, Strzalka K. Lipocalins - a family portrait. J Plant Physiol. 2006; 163:895–915. [PubMed: 16504339]
55. Hieber AD, Bugos RC, Yamamoto HY. Plant lipocalins: violaxanthin de-epoxidase and zeaxanthin epoxidase. Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology. 2000; 1482:84–91.
56. Skerra A. Lipocalins as a scaffold. Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology. 2000; 1482:337–350.
57. Huber R, Schneider MJ, Mayr I, Müller R, Deutzmann R, Suter F, Zuber H, Falk H, Kayser H. Molecular structure of the bilin binding protein (BBP) from *Pieris brassicae* after refinement at 2.0 Å resolution. J Molec Biol. 1987; 198:499–513. [PubMed: 3430616]
58. Huber R, Schneidera M, Eppa O, Mayra I, Messerschmidta A, Pflugratha J, Kayser H. Crystallization, crystal-structure analysis and preliminary molecularmodel of the bilin binding protein from the insect *Pieris brassicae*. J Molec Biol. 1987; 195:423–434. [PubMed: 3656419]
59. Andersen JF, Weichsel A, Balfour CA, Champagne DE, Montfort WR. The crystal structure of nitrophorin 4 at 1.5 angstrom resolution: transport of nitric oxide by a lipocalin-based heme protein. Structure. 1998; 6:1315–1327. [PubMed: 9782054]
60. Flower DR, North ACT, Sansom CE. The lipocalin protein family: structural and sequence overview. Biochim et Biophys Acta. 2000; 1482:9–24.
61. Newcomer ME, Ong DE. Plasma retinol binding protein: structure and function of the prototypic lipocalin. Biochim et Biophys Acta. 2000; 1482:57–64.
62. Kumagai A, Ando R, Miyatake H, Greimel P, Kobayashi T, Hirabayashi Y, Shimogori T, Miyawaki A. A bilirubin-inducible fluorescent protein from eel muscle. Cell. 2013; 153:1602–1611. [PubMed: 23768684]
63. Montelione GT, Anderson S. Structural genomics: Keystone for a human proteome project. Nature Struct Biol. 1999; 6:11–12. [PubMed: 9886282]
64. Liu J, Montelione GT, Rost B. Novel leverage of structural genomics. Nature Biotech. 2007; 25:849–851.
65. Holm L, Rosenström P. Dali server: conservation mapping in 3D. Nucleic Acids Research. 2010; 38:W545–W549. [PubMed: 20457744]
66. Lundell DJ, Glazer AN. Allophycocyanin B: a common β subunit in *Synechococcus* allophycocyanin B (λ_{max} 670 nm) and allophycocyanin (λ_{max} 650 nm). J Biol Chem. 1981; 246:12600–12606. [PubMed: 6795210]
67. Ashby MK, Mullineaux CW. The role of ApcD and ApcF in energy transfer from phycobilisomes to PSI and PSII in a cyanobacterium. Photosynth Res. 1999; 61:169–179.
68. Dong CX, Tang AH, Zhao JD, Mullineaux CW, Shen GZ, Bryant DA. ApcD is necessary for efficient energy transfer from phycobilisomes to photosystem I and helps to prevent photoinhibition in the cyanobacterium *Synechococcus* sp PCC 7002. Biochim Biophys Acta-Bioenerg. 2009; 1787:1122–1128.
69. Zhao J, Shen G, Bryant DA. Photosystem stoichiometry and state transitions in a mutant of the cyanobacterium *Synechococcus* sp. PCC 7002 lacking phycocyanin. Biochim Biosphys Acta. 2001; 1505:248–257.

70. Zhao, J.; Zhou, J.; Bryant, DA. Energy transfer processes in phycobilisomes as deduced from analyses of mutants of *Synechococcus* sp. PCC 7002. In: Murata, N., editor. Research in Photosynthesis. Kluwer; Dordrecht: 1992. p. 25-32.
71. Gindt YM, Zhou JH, Bryant DA, Sauer K. Spectroscopic studies of phycobilisome subcore preparations lacking key core chromophores - Assignment of excited-state energies to the L_{CM}, beta^{18} and alpha^{AP-B} chromophores. Biochim Biophys Acta-Bioenerg. 1994; 1186:153–162.
72. Morimoto K, Sato S, Tabata S, Nakai M. A HEAT-repeats containing protein, IaiH, stablizes the iron-sulfur cluster bound to the cyanobacterial IscA homologue, IscA2. J Biochem (Tokyo). 2003; 134:211–217. [PubMed: 12966069]
73. Morimoto K, Nishio K, Nakai M. Identification of a novel prokaryotic HEAT-repeats-containing protein which interacts with a cyanobacterial IscA homolog. FEBS Lett. 2002; 519:123–127. [PubMed: 12023030]
74. Takano H, Gusella J. The predominantly HEAT-like motif structure of huntingtin and its association and coincident nuclear entry with dorsal, an NF-kB/Rel/dorsal family transcription factor. BMC Neurosci. 2002; 3:15. [PubMed: 12379151]
75. Andrade MA, Petosa C, O'Donoghue SI, Müller CW, Bork P. Comparison of ARM and HEAT protein repeats. J Mol Biol. 2001; 309:1–18. [PubMed: 11491282]
76. Kelley LA, Sternberg MJE. Protein structure prediction on the web: a case study using the Phyre server. Nature Protocols. 2009; 4:363–371.
77. Kupka M, Zhang J, Fu WL, Tu JM, Bohm S, Su P, Chen Y, Zhou M, Scheer H, Zhao KH. Catalytic mechanism of S-type phycobiliprotein lyase chaperone-like action and functional amino acid residues. J Biol Chem. 2009; 284:36405–36414. [PubMed: 19864423]
78. Arciero DM, Bryant DA, Glazer AN. *In vitro* attachment of bilins to apophycocyanin I: specific covalent adduct formation at cysteinyl residues involved in phycocyanobilin binding in C-phycocyanin. J Biol Chem. 1988; 263:18343–18349. [PubMed: 3142876]
79. Arciero DM, Dallas JL, Glazer AN. *In vitro* attachment of bilins to apophycocyanin: II: determination of the structures of tryptic bilin peptides derived from the phycocyanobilin adduct. J Biol Chem. 1988; 263:18350–18357. [PubMed: 3192538]
80. Schluchter, WM.; Glazer, AN. Biosynthesis of phycobiliproteins in cyanobacteria. In: Peschek, GA.; Löffelhardt, W.; Schmetterer, G., editors. The Phototrophic Prokaryotes. Plenum; New York: 1999. p. 83-95.
81. Bohm S, Endres S, Scheer H, Zhao KH. Biliprotein chromophore attachment - Chaperone-like function of the PecE subunit of alpha-phycoerythrocyanin lyase. J Biol Chem. 2007; 282:25357–25366. [PubMed: 17595164]
82. Wu S-H, Lagarias JC. Defining the bilin lyase domain: Lessons from the extended phytochrome superfamily. Biochemistry. 2000; 39:13487–13495. [PubMed: 11063585]
83. Tu JM, Kupka M, Boehm S, Ploescher M, Eichacker L, Zhao KH, Scheer H. Intermediate binding of phycocyanobilin to the lyase, CpeS1, and transfer to apoprotein. Photosynth Res. 2008; 95:163–168. [PubMed: 17912606]
84. Schulz H, Hennecke H, Thöny-Meyer L. Prototype of a heme chaperone essential for cytochrome *c* maturation. Science. 1998; 281:1197–1200. [PubMed: 9712585]
85. Alvey RM, Biswas A, Schluchter WM, Bryant DA. Effects of Modified Phycobilin Biosynthesis in the Cyanobacterium *Synechococcus* sp Strain PCC 7002. J Bacteriol. 2011; 193:1663–1671. [PubMed: 21296968]
86. Miller CA, Leonard HS, Pinsky IG, Turner BM, Williams SR, Harrison L Jr, Fletcher AF, Shen G, Bryant DA, Schluchter WM. Biogenesis of phycobiliproteins.III. CpcM is the asparagine methyltransferase for phycobiliprotein β-subunits in cyanobacteria. J Biol Chem. 2008; 283:19293–19300. [PubMed: 18482977]
87. Dammeyer T, Bagby SC, Sullivan MB, Chisholm SW, Frankenberg-Dinkel N. Efficient phage-mediated pigment biosynthesis in oceanic cyanabacteria. Curr Biol. 2008; 18:442–448. [PubMed: 18356052]
88. Drenth, J. Principles of Protein X-Ray Crystallography. 2. Springer; NY: 1999. edition ed
89. Notredame C, Higgins D, Heringa J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. J Mol Biol. 2000; 302:205–217. [PubMed: 10964570]

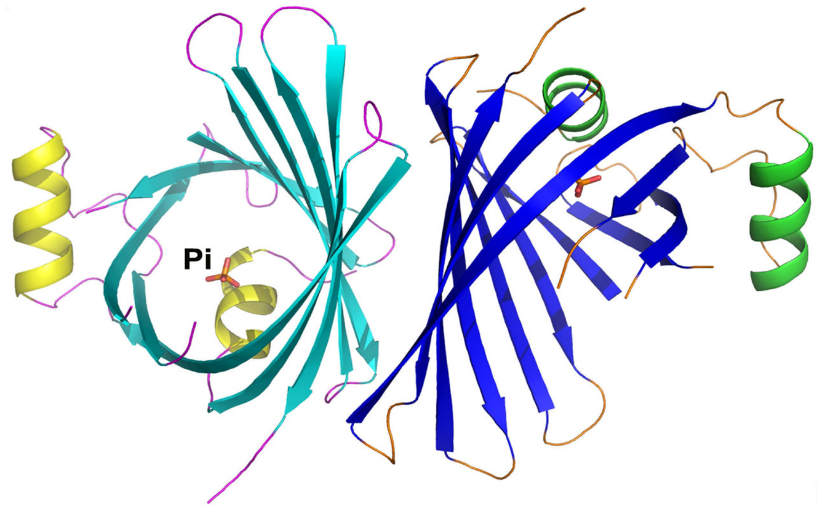

Fig. 1.

Structure of *Te*CpcS from *Thermosynechococcus elongatus* BP-1 (PDB id: 3BDR). The crystal structure of the homodimer *Te*CpcS. The two α-helices, 10 β-strands, and the associated loops of each subunit are depicted in yellow/cyan/magenta for subunit A and in green/blue/orange for subunit B, respectively. A phosphate ion co-crystallized with each subunit is shown as ball-and-stick model.

NIH-PA Author Manuscript
NIH-PA Author Manuscript

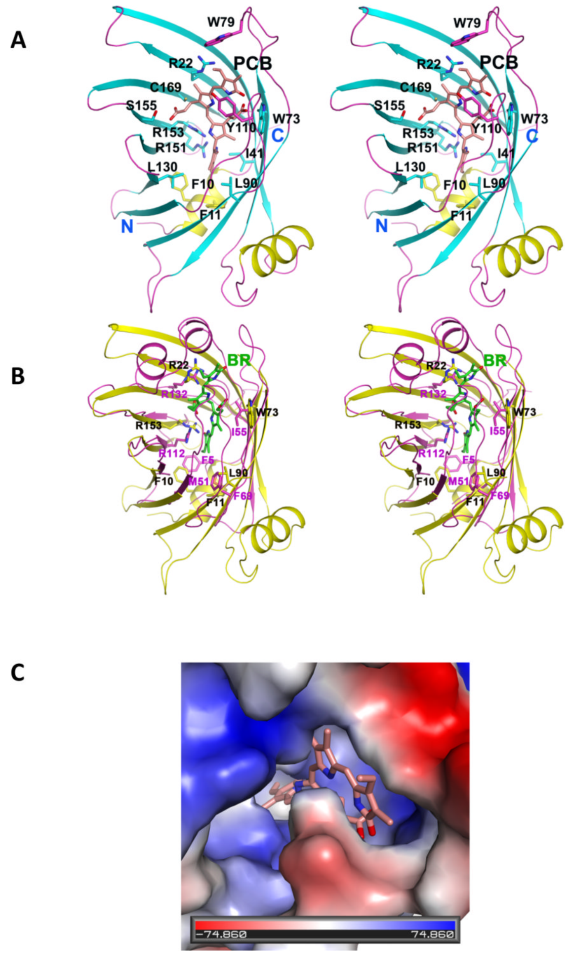

Fig. 2.

Structures of *Te*CpcS modeled with PCB. **(A)** Stereo pair view of a *Te*CpcS protomer with PCB modeled in the β-barrel. The modeled PCB and the side chains of strictly conserved and conservatively substituted residues are shown in ball-and-stick models. The two loops comprising residues 77–83 and 108–117 which are disordered in the *Te*CpcS structure are also modeled. The N- and C-termini of Te-CpcS are labeled. **(B)** Stereo pair view of the structural superposition of UnaG bound to bilirubin (PDB id: 4I3B; in magenta) and *Te*CpcS (in yellow). The side chains of identical and similar residues, and the bilirubin (BR) bound to UnaG are depicted as stick models and are labeled. **(C)** View of a PCB model docked into the cavity of *Te*CpcS, which is shown with a surface charged map representation. The stick model of the docked PCB is shown in pink oriented such that the vinyl group on Ring A is close to the surface and Ring D is buried in the cavity/barrel, while the *Te*CpcS surface is depicted in color with grey/white representing neutral/hydrophobic amino acids, blue representing positively charged amino acids, and red representing negatively charged amino acids.

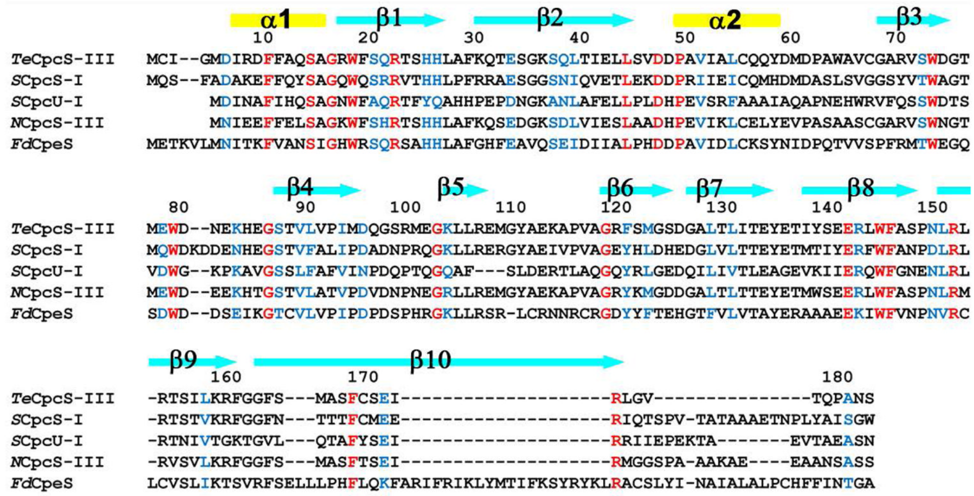

Fig. 3.

Sequence alignment of representative CpcS-type proteins performed using T-COFFEE [89]. For each sequence, the organism name appears as italic letters before the protein: *Thermosynechococcus elongatus* BP-1 (*Te*), *Synechococcus* sp. PCC 7002 (*S*), *Nostoc sp.* PCC 7120 (*N*), and *Fremyella diplosiphon* (*Fd*). Identical and conservatively replaced residues in all sequences are shown in *red* and *blue*, respectively. Secondary structural elements, observed in the crystal structure of *Te*CpcS, are shown above the alignment with α-helices and β-strands represented by *rectangles* and *arrows*, respectively.

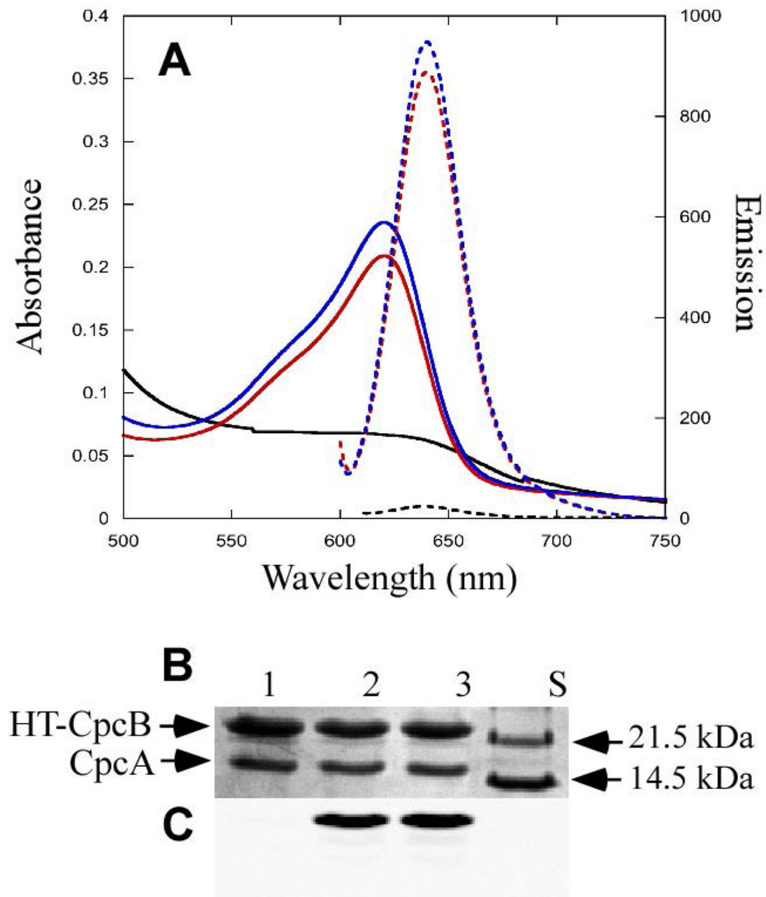

Fig. 4.

Chromophorylation of CpcB by *Te*CpcS versus CpcS/CpcU. (**A)** Absorbance (solid) and fluorescence emission (dashed) spectra of HT-CpcBA purified from recombinant *E. coli* cells containing pCpcBA with pPcyA and expressing either *Te*CpcS (blue), CpcSU (red) or no lyase (black). (**B)** Coomassie-stained SDS-polyacrylamide gel of HT-CpcBA purified from cells containing pCpcBA, pPcyA and expressing either CpcSU (lane 2), *Te*CpcS (lane 3), or no lyase (lane 1). Molecular mass standards were loaded in lane S. (**C)** Zn-enhanced bilin fluorescence (excitation at 635 nm) of the gel in panel B.

NIH-PA Author Manuscript
NIH-PA Author Manuscript

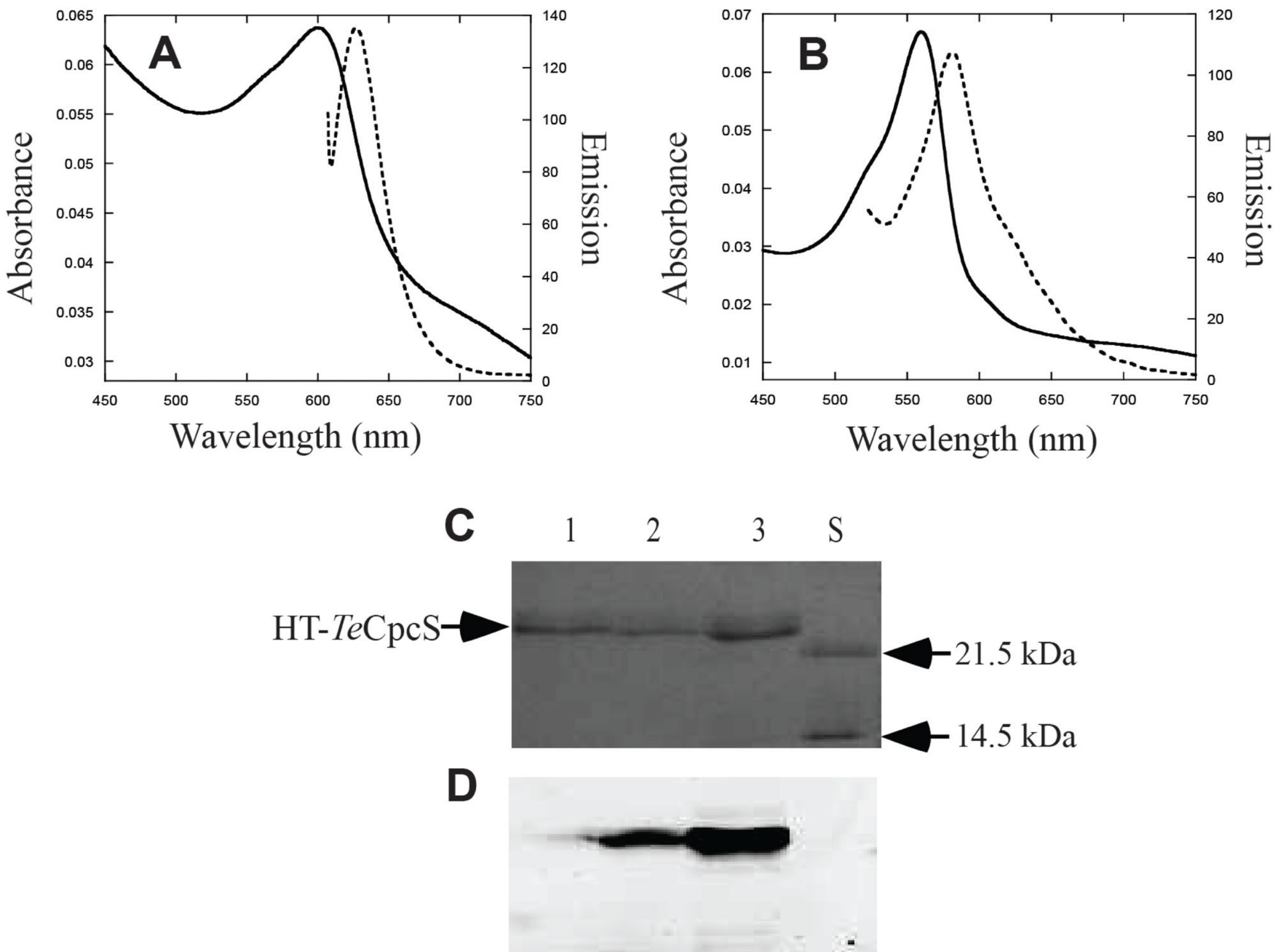

Fig. 5.

Bilin binding to the lyase *Te*CpcS. **(A)** Absorbance (solid) and fluorescence emission (dashed) spectra of HT-*Te*CpcS purified from recombinant *E. coli* cells containing pTER13-21 and pPcyA. **(B)** Absorbance (solid) and fluorescence emission (dashed) spectra of purified HT-*Te*CpcS from cells purified from recombinant *E. coli* containing pTER13-21and pPebS. **(C)** Coomassie-stained SDS-polyacrylamide gel of HT-*Te*CpcS purified from cells containing pTER13-21 alone (lane 1), pTER13-21and pPcyA (lane 2) or pTER13-21and pPebS (lane 3). Molecular mass standards were loaded in lane S. **(D)** Zn-enhanced bilin fluorescence of the gel (excitation at 532 nm) in panel C.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

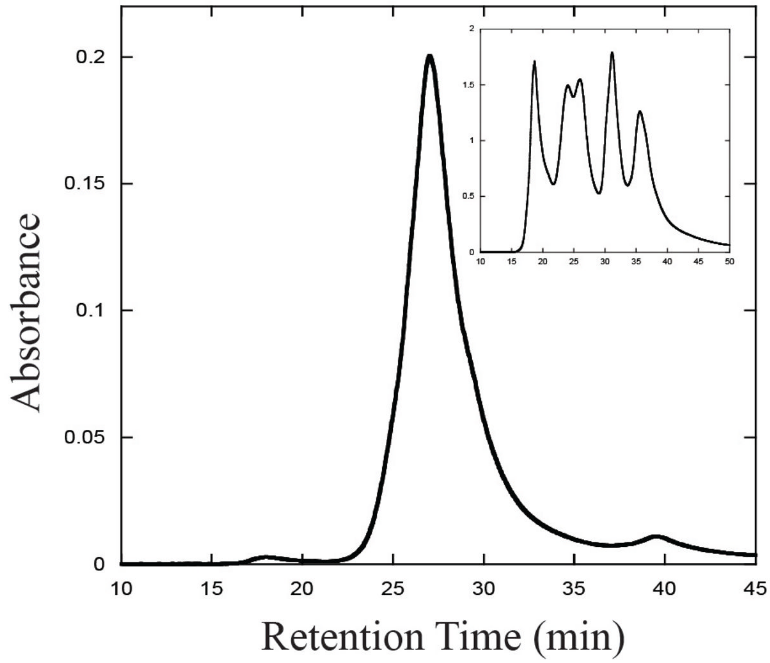

Fig. 6.

Size exclusion chromatography of recombinant *Te*CpcS. Recombinant HT-*Te*CpcS was injected onto the HPLC column and protein eluates were monitored at 280 nm. The calculated M_r of the fraction eluting at 26.9 min was 45,600, which suggests that HT-*Te*CpcS is a dimer under the conditions of this experiment. The inset graph shows the chromatogram for the molecular weight standards.

Table 1

Summary of plasmids used in this study

Plasmid Name	Recombinant proteins produced[a]	Parent vector	Antibiotic[b]	Reference
pApcAB	*Synechococcus* sp. PCC 7002 HT-ApcA and ApcB	pET100	Ap	[18]
pApcDB	*Synechococcus* sp. PCC 7002 HT-ApcD and ApcB	pET100	Ap	[86]
pApcF	*Synechococcus* sp. PCC 7002 HT-ApcF	pET100	Ap	[86]
pApcF/CpcS	*Synechococcus* sp. PCC 7002 HT-ApcF and *Thermosynechococcus elongatus* CpcS	pCOLA Duet	Km	This study
pPcyA	*Synechocystis* sp. PCC 6803 HO1 and *Synechococcus* sp. PCC 7002 HT-PcyA	pACYC Duet	Cm	[13]
pCpcSU	*Synechococcus* sp. PCC 7002 CpcS-I and CpcU	pCOLA Duet	Km	[13]
pCpcBA	*Synechocystis* sp. PCC 6803 HT-CpcB and CpcA	pCDF Duet	Sp	[13]
pPebS	Myovirus HO1 and HT-PebS	pACYC Duet	Cm	[87]
pPebS2	Myovirus HO1 and HT-PebS	pCOLA Duet	Km	This study
p*Te*CpcS	*Thermosynechococcus elongatus* CpcS	pCOLA Duet	Km	This study
pTER13-21	*Thermosynechococcus elongatus* HT-CpcS	pET21c	Ap	This study
pTER13-30	*Thermosynechococcus elongatus* HT-CpcS	pET30c	Km	This study
pTER13(R151G)	*Thermosynechococcus elongatus* HT-CpcS (R151G)	pTER13-30	Km	This study
pTER13(S155G)	*Thermosynechococcus elongatus* HT-CpcS (S155G)	pTER13-30	Km	This study
pTER13(C2S)	*Thermosynechococcus elongatus* HT-CpcS (C2S)	pTER13-30	Km	This study
pTER13(C169S)	*Thermosynechococcus elongatus* HT-CpcS (C169S)	pTER13-30	Km	This study
pTER13(C2S/C169S)	*Thermosynechococcus elongatus* HT-CpcS (C2S/C169S)	pTER13-30	Km	This study
pHy2	*Synechocystis* sp. PCC 6803 HO1 and *Arabidopsis thaliana* HT-Hy2	pACYC Duet	Cm	[24]

[a] Proteins that would be produced as fusions are indicated as HT; meaning His tagged.

[b] Antibiotic resistance used to select for the presence of the plasmid (Ap: ampicillin; Cm: chloramphenicol; Km: kanamycin; Sp: spectinomycin)

Table 2

X-ray data-quality and structure-refinement statistics [a]

Data collection	
Space group	$P4_12_12$
Cell dimensions	
a, *b*, *c* (Å)	75.014, 75.014, 83.359
α, β, γ (°)	90, 90, 90
Wavelength (Å)	0.979
Resolution range (Å)	50.0-2.80 (2.90-2.80)*
R_{sym} or R_{merge}	0.081 or 0.059 (0.691 or 0.634)
Total reflections	362,419
Observed reflections	11,252
$I/\sigma I$	28.2 (4.6)
Completeness (%)	99.9 (100.0)
Redundancy	10.1 (9.7)
Refinement	
Resolution (Å)	20.0-2.80
No. reflections	10,246
R_{work}/R_{free}	0.236/0.262 (0.312/0.395)
No. atoms	1231
Protein	1225 (subunit A: 4–76, 84–107, 118–176)
Ligand	5
Water	1
B-factors (Å^2)	
Protein	66.5
Ligand/ion	97
Water	56.8
R.m.s. deviations	
Bond lengths (Å)	0.008
Bond angles (°)	1.40
Ramachandran Distribution (%)	
Favored	86.6
Allowed	13.4
Generously allowed	0.0
Disallowed	0.0
PDB id	3BDR

[a] Standard parameter definitions were used [88].

* Highest-resolution shell is shown in parentheses.

Statistics belong to one crystal.

Table 3

Spectral properties for PC and AP subunits chromophorylated with PCB, PEB or PΦB

Holo recombinant PBPs (Plasmid present)	λ_{max} (nm)($Q_{Vis/UV}$)	Fluorescence Emission λ_{max} (nm)
HT-CpcB (pCpcBA + pPcyA)[a]	620/394 (4.3)	644
HT-CpcB (pCpcBA + pPcyA)[b]	621/393 (4.7)	644
HT-CpcB (pCpcBA + pPebS)[a]	557/372 (4.95)	573
HT-CpcB (pCpcBA + pHY2)[a]	635/347 (2.25)	647
HT-ApcA/ApcB (pApcAB + pPcyA)[a]	614/392 (5.3)	632
HT-ApcA/ApcB (pApcAB + pPebS)[a]	560/376 (8.3)	571
HT-ApcA/ApcB (pApcAB + pHY2)[a]	629/391 (4.8)	648
HT-ApcD (pApcDB + pPcyA)[a]	615, 672[c]/370 (1.2)	635, 675
HT-ApcD (pApcDB + pPebS)[a]	572/371 (2.8)	571
HT-ApcD (pApcDB + pHY2)[a]	629/391 (2.4)	630
HT-ApcF (pApcF + pPcyA)[a]	615/393 (4.2)	632
HT-ApcF (pApcF + pPebS)[a]	560/376 (8.1)	572
HT-ApcF (pApcF + pHY2)[a]	630/387 (3.5)	648

[a]The construct was coexpressed with p*Te*CpcS

[b]The construct was coexpressed with pCpcSU

[c]Indicates a second peak